Optimality conditions. KKT

Background

Extreme value (Weierstrass) theorem

Let $S \subset \mathbb{R}^n$ be compact set and $f(x)$ continuous function on S. So that, the point of the global minimum of the function $f(x)$ on S exists.

Lagrange multipliers

Consider simple yet practical case of equality constraints:

$f(x) = (x-5)^2$
$|x| = 3$

$$
\begin{split}
& f(x) \to \min\limits_{x \in \mathbb{R}^n} \\
\text{s.t. } & h_i(x) = 0, i = 1, \ldots, p
\end{split}
$$

The basic idea of Lagrange method implies switch from conditional to unconditional optimization through increasing the dimensionality of the problem: множ. Лагр.

$$
\boxed{L(x, \nu) = f(x) + \sum\limits_{i=1}^p \nu_i h_i(x) \to \min\limits_{x \in \mathbb{R}^n, \nu \in \mathbb{R}^p}}
$$

General formulations and conditions

$$
\boxed{f(x) \to \min\limits_{x \in S}}
$$

бюджетное мн-во

We say that the problem has a solution if the budget set **is not empty**: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

Optimization on the general set S.

Direction $d \in \mathbb{R}^n$ is a feasible direction at $x^* \in S \subseteq \mathbb{R}^n$ if small steps along d do not take us outside of S.

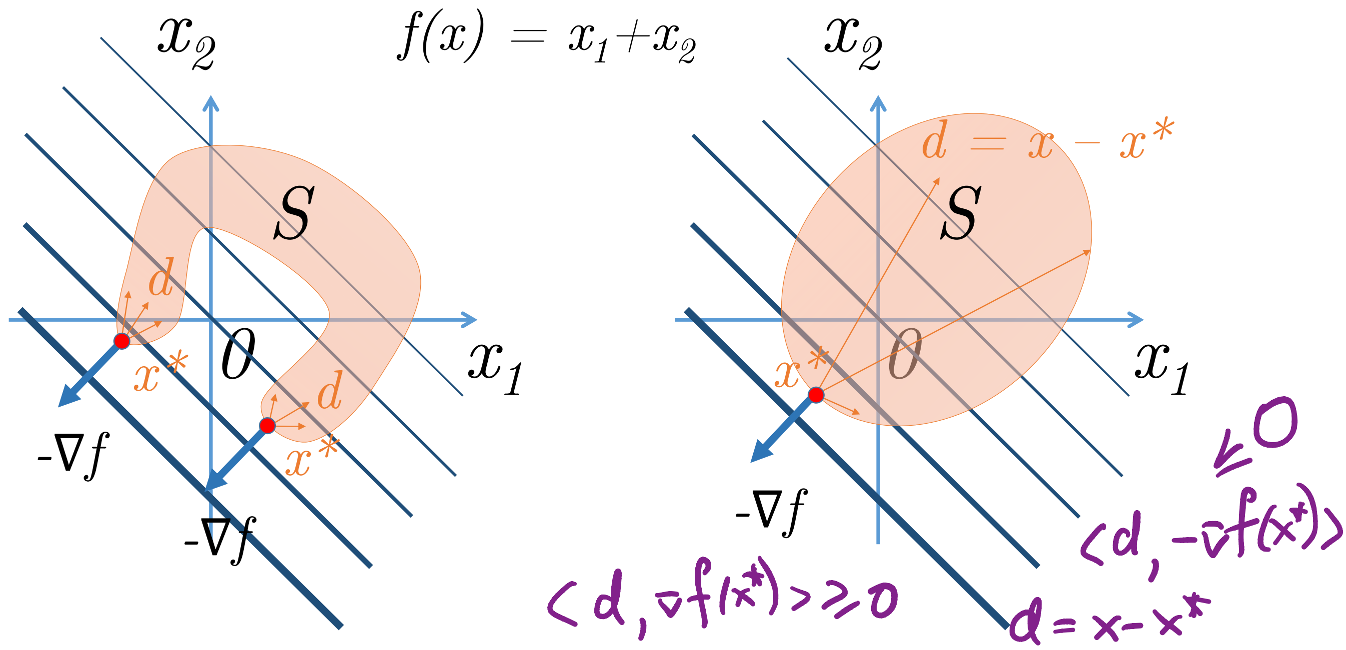
Consider a set $S \subseteq \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that $x^* \in S$ is a point of local minimum for f over S, and further assume that f is continuously differentiable around x^*.

1. Then for every feasible direction $d \in \mathbb{R}^n$ at x^* it holds that $\nabla f(x^*)^\top d \geq 0$

x^* – решение зад.

2. If, additionally, S is convex then

$$\boxed{\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in S.}$$



Unconstrained optimization

General case

Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function.

$$\boxed{f(x) \to \min_{x \in \mathbb{R}^n}} \tag{UP}$$

If x^* - is a local minimum of $f(x)$, then:

$$\boxed{\nabla f(x^*) = 0} \tag{UP:Nec.}$$

Необх.

If $f(x)$ at some point x^* satisfies the following conditions:

$$\boxed{H_f(x^*) = \nabla^2 f(x^*) \succ (\prec) 0,} \tag{UP:Suff.}$$

дост.

then (if necessary condition is also satisfied) x^* is a local minimum(maximum) of $f(x)$.

Note, that if $\nabla f(x^*) = 0, \nabla^2 f(x^*) = 0$, i.e. the hessian is positive *semidefinite*, we cannot be sure if x^* is a local minimum (see Peano surface $f(x, y) = (2x^2 - y)(y - x^2)$).

$\min_{x \in S} f(x)$ – вып.

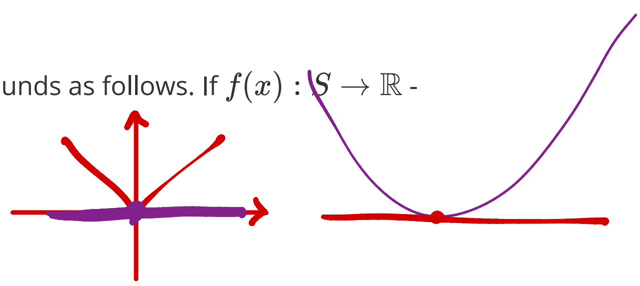
Convex case

It should be mentioned, that in **convex** case (i.e., $f(x)$ is convex) necessary condition becomes sufficient. Moreover, we can generalize this result on the class of non-differentiable convex functions.

Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ - convex function, then the point x^* is the solution of (UP) if and only if:

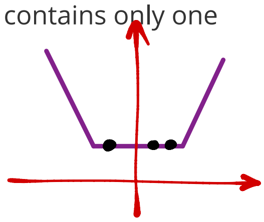
$$\boxed{0_n \in \partial f(x^*)}$$

One more important result for convex constrained case sounds as follows. If $f(x) : S \to \mathbb{R}$ - convex function defined on the convex set S, then:



- Any local minima is the global one.
- The set of the local minimizers S^* is convex.

- If $f(x)$ - strictly or strongly (different cases 😀) convex function, then S^* contains only one single point $S^* = x^*$.



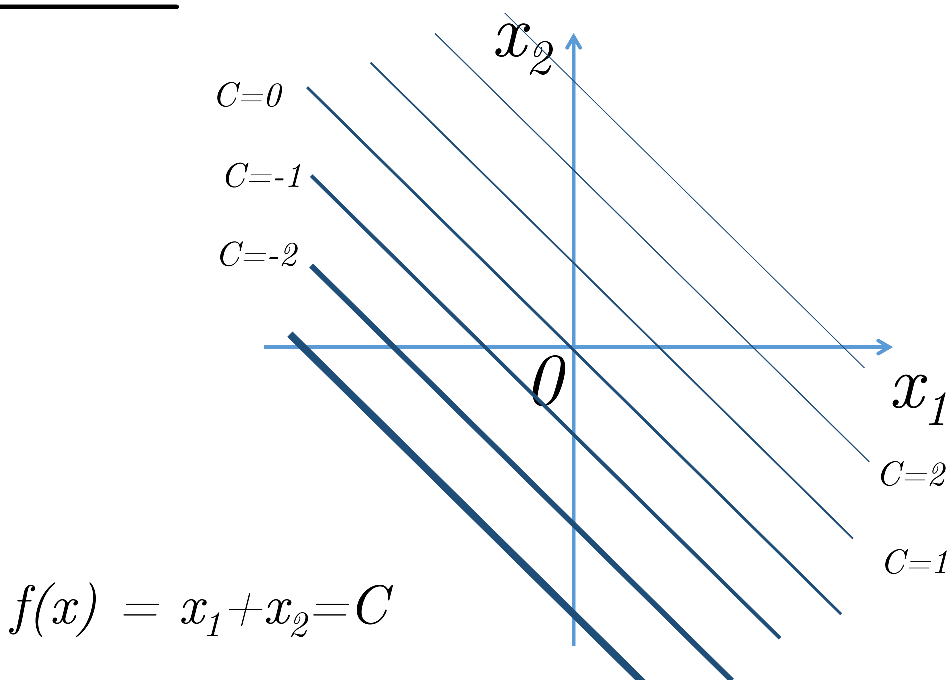
Optimization with equality conditions

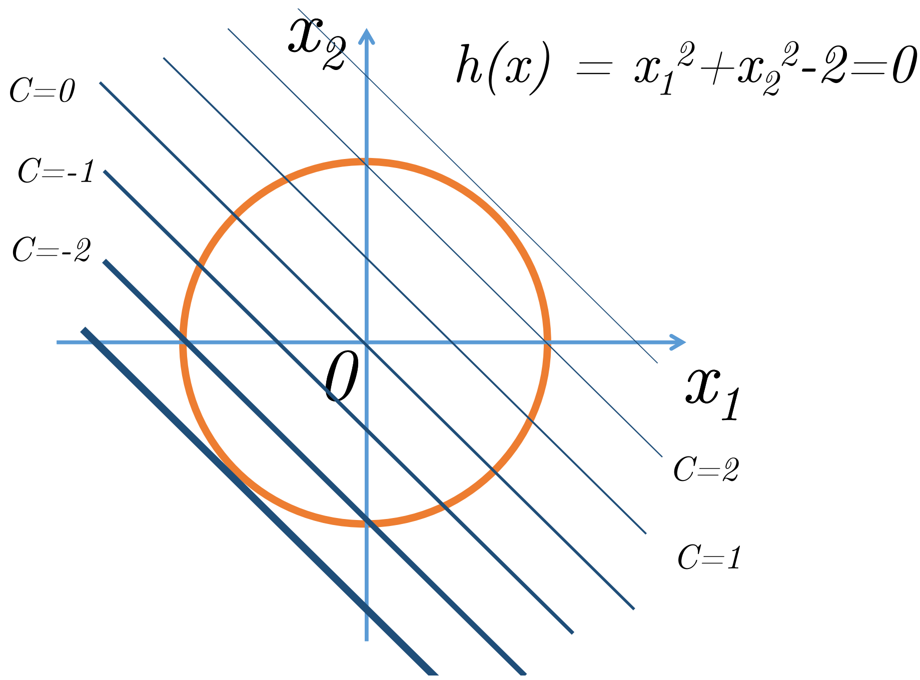
Intuition

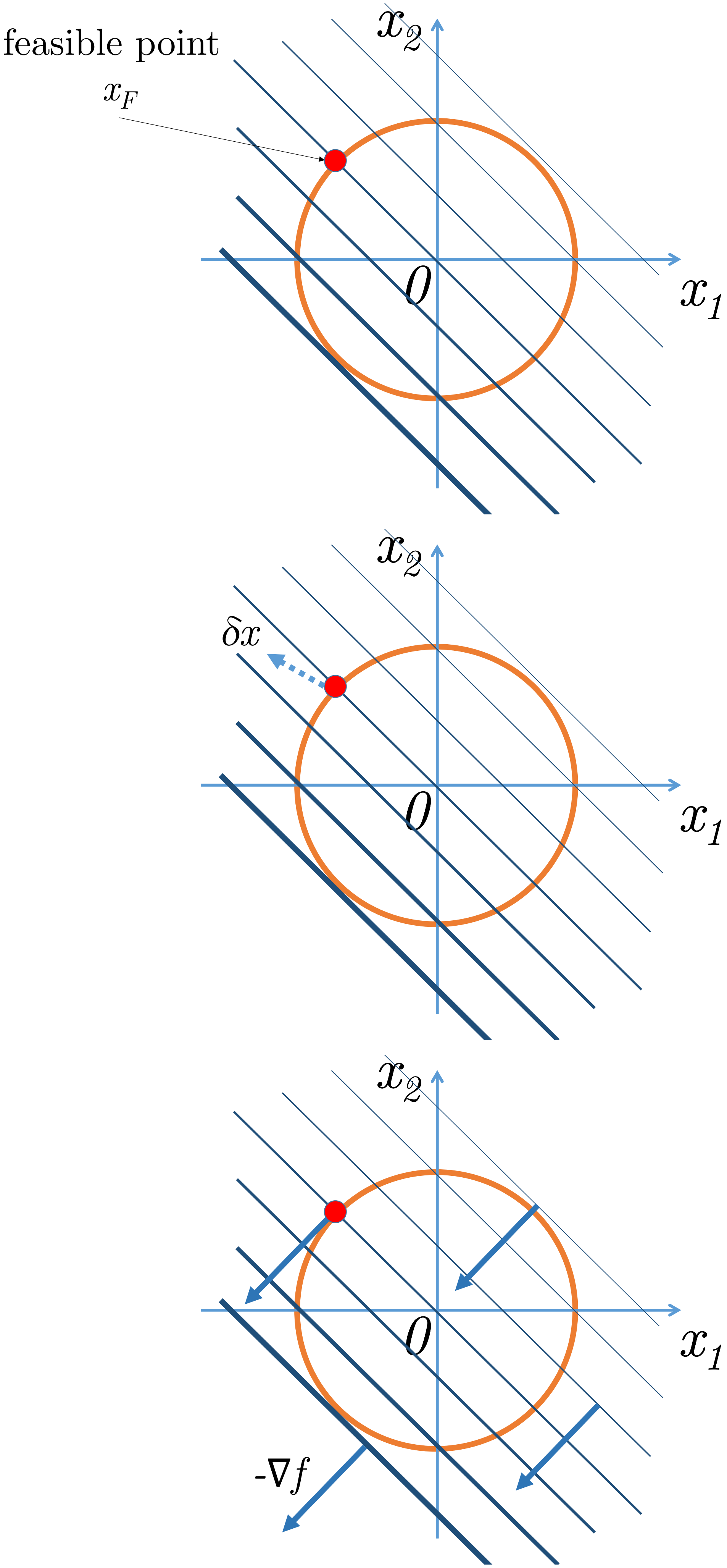
Things are pretty simple and intuitive in unconstrained problem. In this section we will add one equality constraint, i.e.

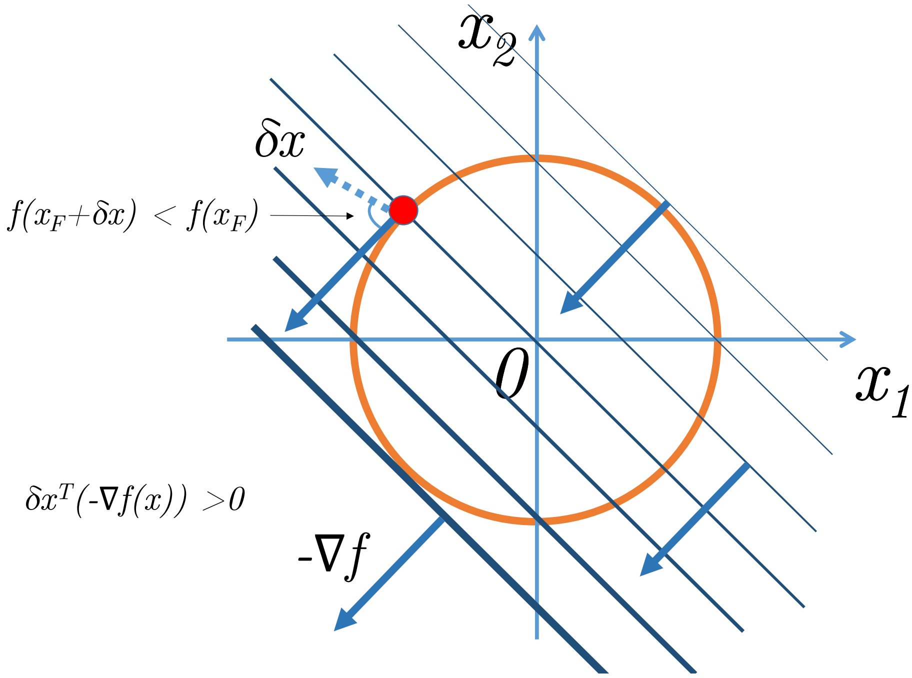
$$
\boxed{
\begin{aligned}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
& \text{s.t. } h(x) = 0
\end{aligned}
}
$$

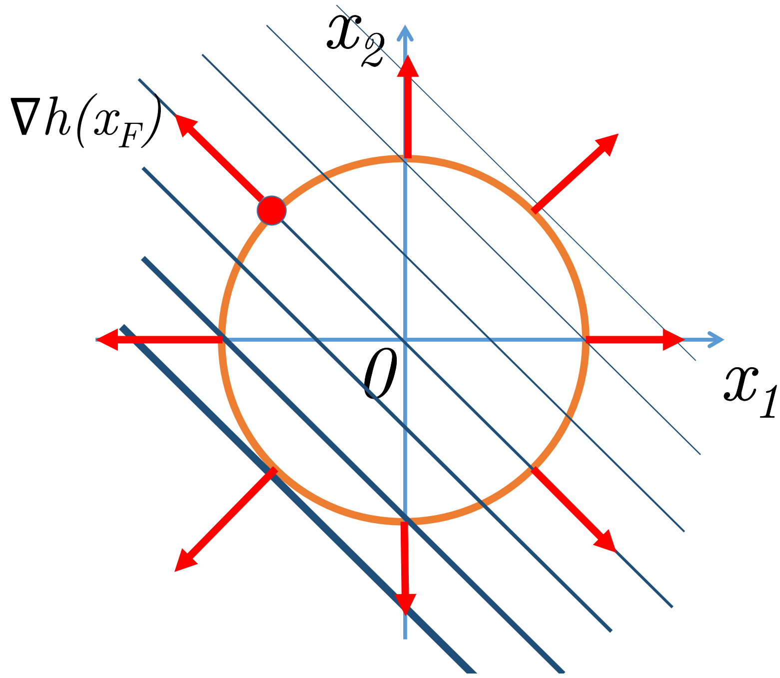
We will try to illustrate approach to solve this problem through the simple example with $f(x) = x_1 + x_2$ and $h(x) = x_1^2 + x_2^2 - 2$

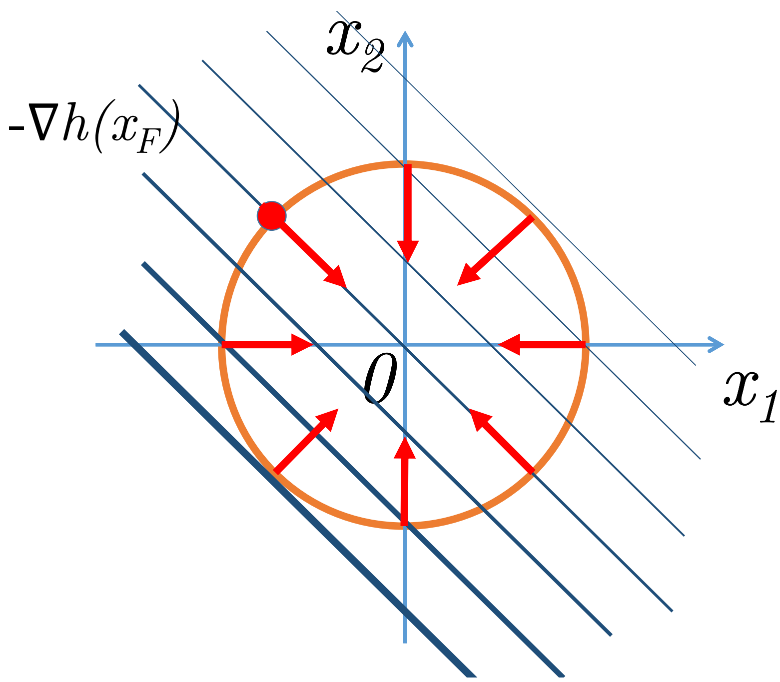





feasible point
x_F
x_2
x_1
0
δx
x_2
x_1
0
x_2
x_1
0
$-\nabla f$


x_2
δx
$f(x_F+\delta x) < f(x_F)$
0
x_1
$\delta x^T(-\nabla f(x)) > 0$
$-\nabla f$


x_2
$\nabla h(x_F)$
0
x_1


x_2
$-\nabla h(x_F)$
0
x_1

Generally: in order to move from x_F along the budget set towards decreasing the function, we need to guarantee two conditions:

$$\boxed{\langle \delta x, \nabla h(x_F) \rangle = 0}$$

$$\boxed{\langle \delta x, -\nabla f(x_F) \rangle > 0}$$

НАХ. В БЮДЖ. МН.

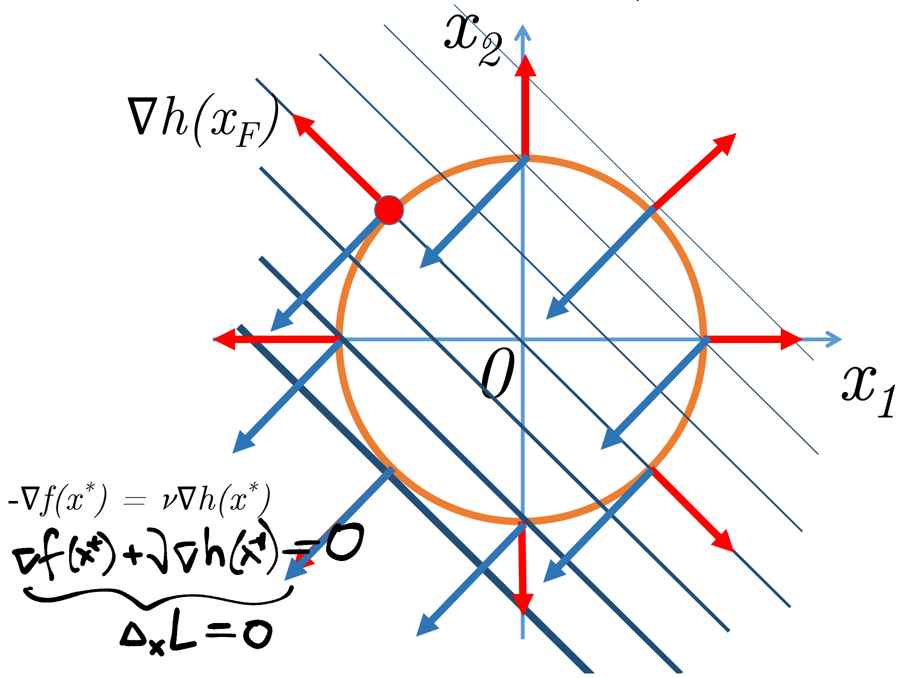
$f \downarrow\downarrow$

Let's assume, that in the process of such a movement we have come to the point where

$$-\nabla f(x) = \nu \nabla h(x)$$

$$\langle \delta x, -\nabla f(x) \rangle = \langle \delta x, \nu \nabla h(x) \rangle = 0$$

Then we came to the point of the budget set, moving from which it will not be possible to reduce our function. This is the local minimum in the constrained problem :)



So let's define a Lagrange function (just for our convenience):

$$\boxed{L(x, \nu) = f(x) + \nu h(x)}$$

Then the point x^* be the local minimum of the problem described above, if and only if:

$$
\begin{split}
& \text{Necessary conditions} \\
1)\ \nabla L = 0 \quad & \boxed{\nabla_x L(x^*, \nu^*) = 0} \text{ that's written above} \\
& \boxed{\nabla_\nu L(x^*, \nu^*) = 0} \text{ budget constraint} \quad h(x) = 0 \\
& \text{Sufficient conditions} \\
2) \quad & \langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle \geq 0, \\
& \forall y \neq 0 \in \mathbb{R}^n : \nabla h(x^*)^\top y = 0
\end{split}
$$

We should notice that $L(x^*, \nu^*) = f(x^*)$.

General formulation

$$
\begin{split}
& f(x) \to \min\limits_{x \in \mathbb{R}^n} \\
\text{s.t. } & h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
\tag{ECP}
$$

Solution

В случае $x \in S$

$$
\boxed{L(x, \nu) = f(x) + \sum_{i=1}^{p} \nu_i h_i(x) = f(x) + \nu^\top h(x)}
$$

$L(x, \nu) = f(x)$

Let $f(x)$ and $h_i(x)$ be twice differentiable at the point x^* and continuously differentiable in some neighborhood x^*. The local minimum conditions for $x \in \mathbb{R}^n, \nu \in \mathbb{R}^m$ are written as

ECP: Necessary conditions

$\nabla_x L(x^*, \nu^*) = 0$ n ур-ий

$\nabla_\nu L(x^*, \nu^*) = 0$ p ур-ий

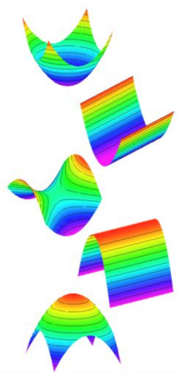
ECP: Sufficient conditions

$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$

$\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y = 0$

Depending on the behavior of the Hessian, the critical points can have a different character.

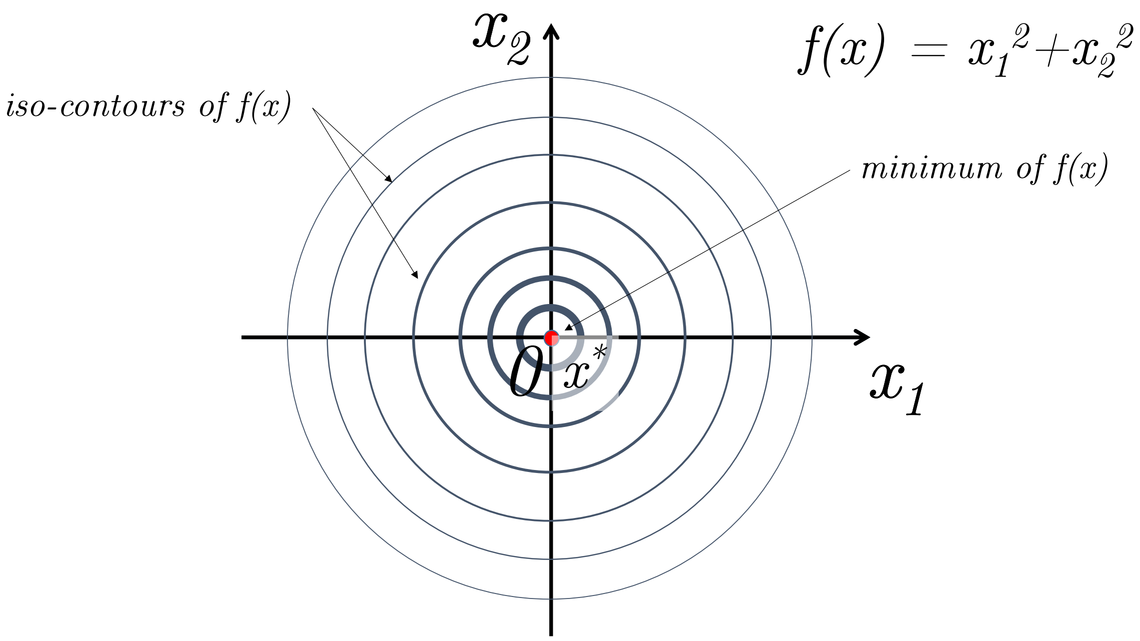
$\mathbf{y}^T\mathbf{H}\mathbf{y}$	λ_i	**Definiteness H**	**Nature x***
>0		Positive d.	Minimum
≥ 0		Positive semi-d.	Valley
$\neq 0$		Indefinite	Saddlepoint
≤ 0		Negative semi-d.	Ridge
<0		Negative d.	Maximum

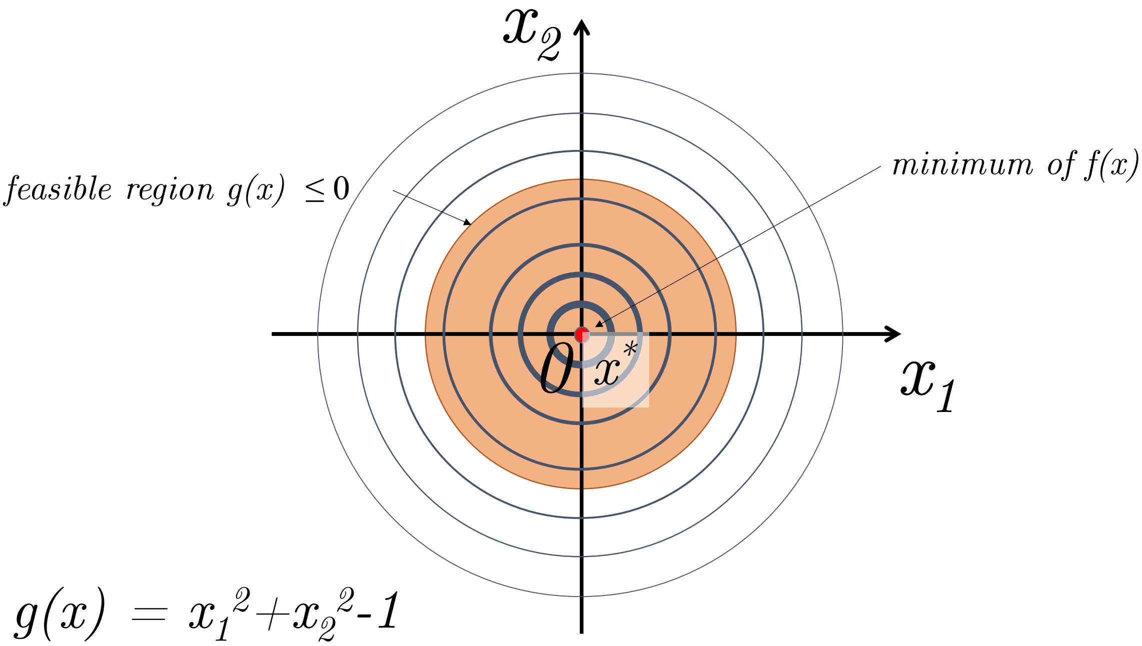


Optimization with inequality conditions

Example

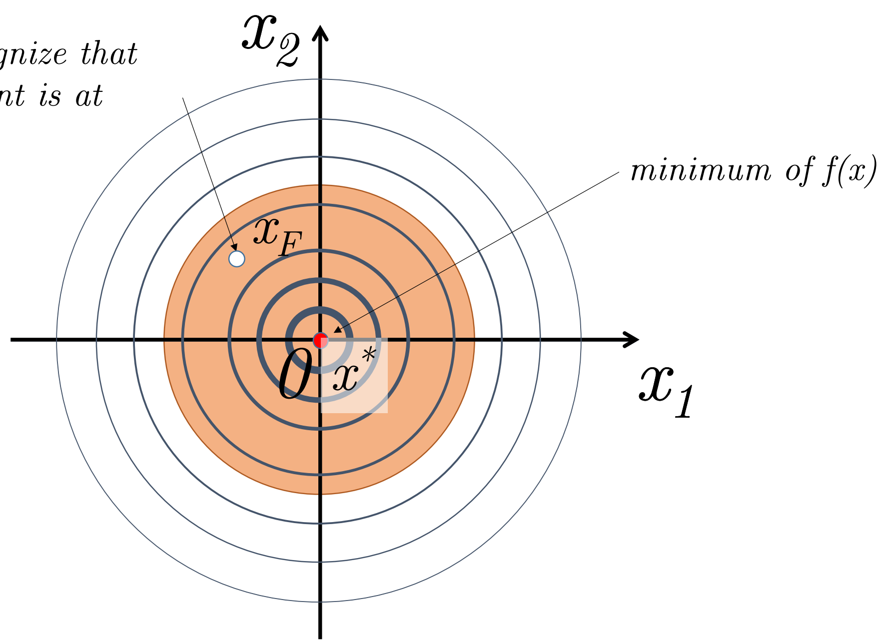
$$
\boxed{\begin{gathered} f(x) = x_1^2 + x_2^2 \quad g(x) = x_1^2 + x_2^2 - 1 \\ f(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) \leq 0 \end{gathered}}
$$



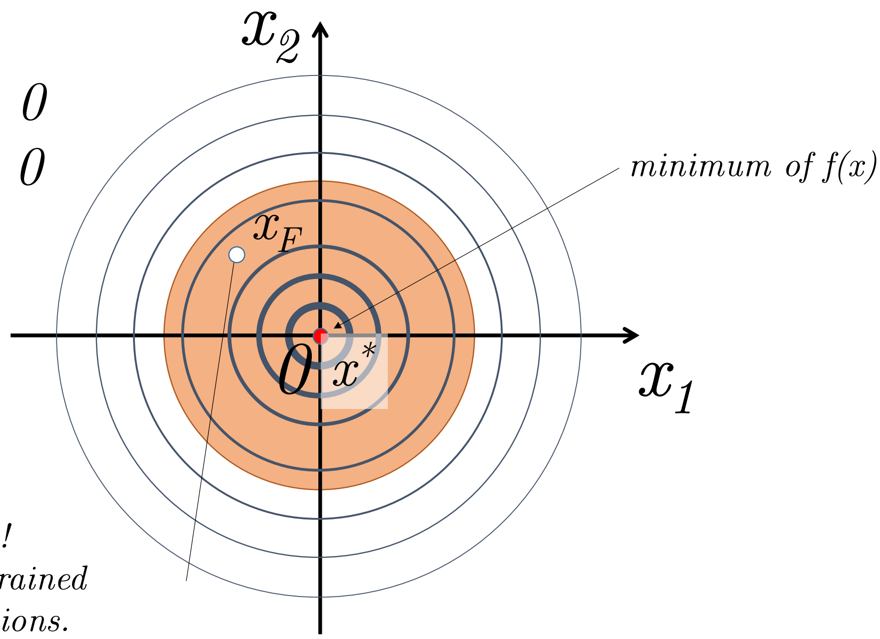


$$g(x) = x_1^2 + x_2^2 - 1$$

How can we recognize that some feasible point is at local minimum?



$$\nabla f(x_F) = 0$$
$$\nabla^2 f(x_F) \succ 0$$



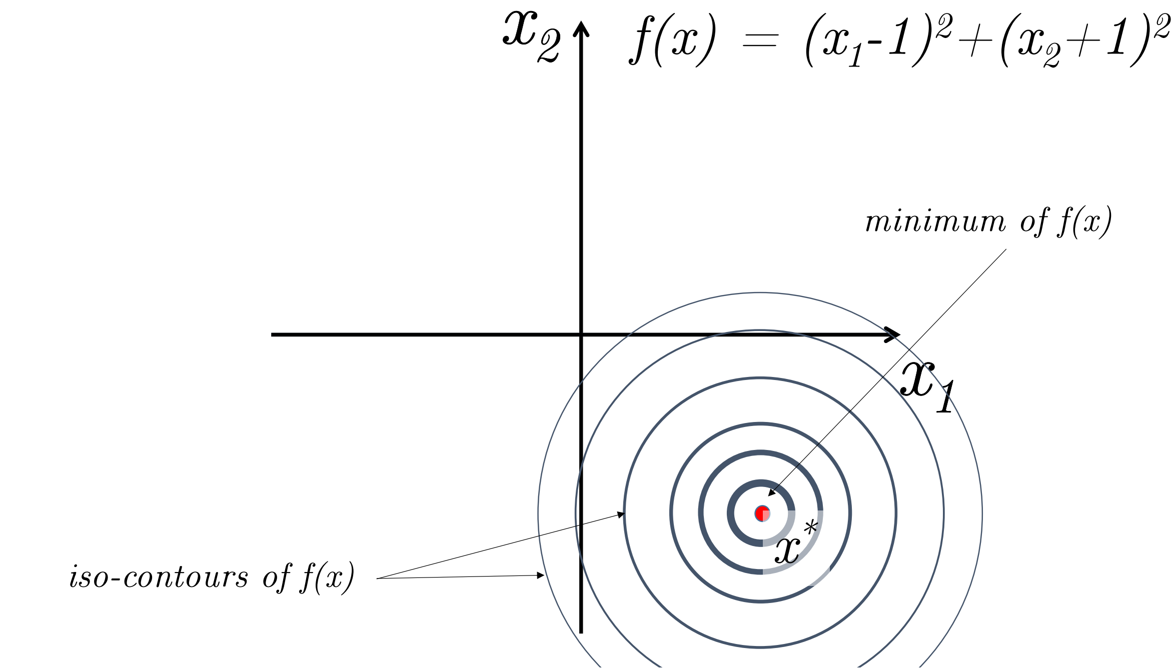
Easy in this case!
Just use unconstrained optimality conditions.

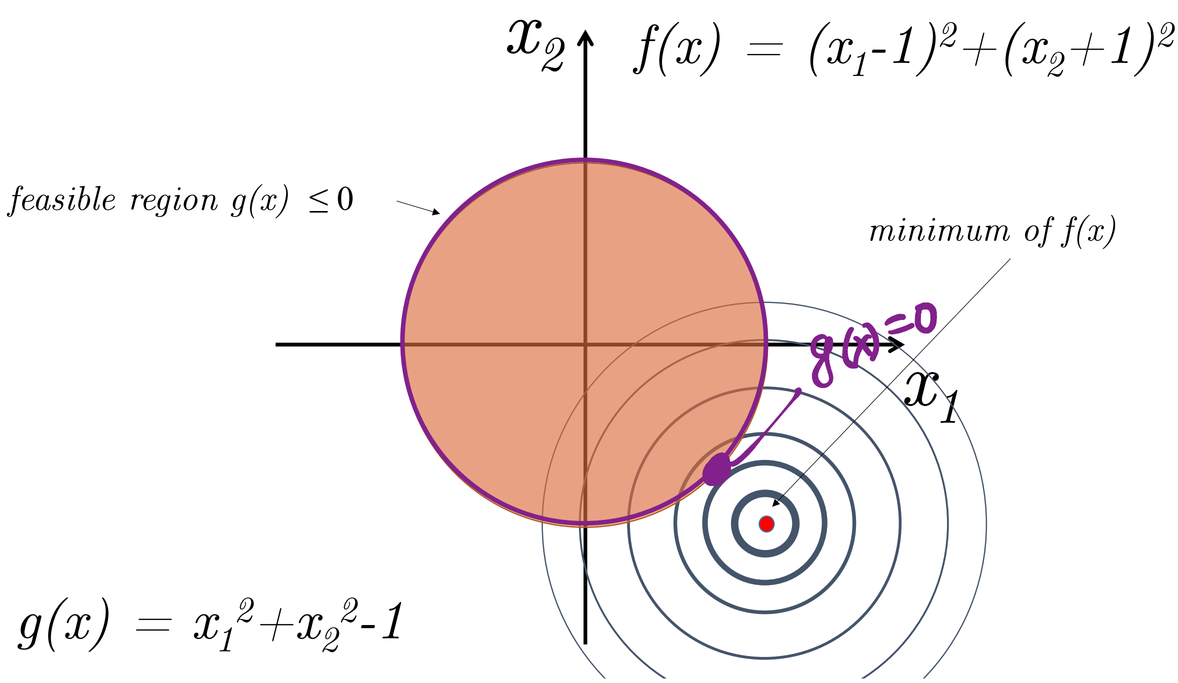
Thus, if the constraints of the type of inequalities are inactive in the constrained problem, then don't worry and write out the solution to the unconstrained problem. However, this is not the whole story 🤔. Consider the second childish example

$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

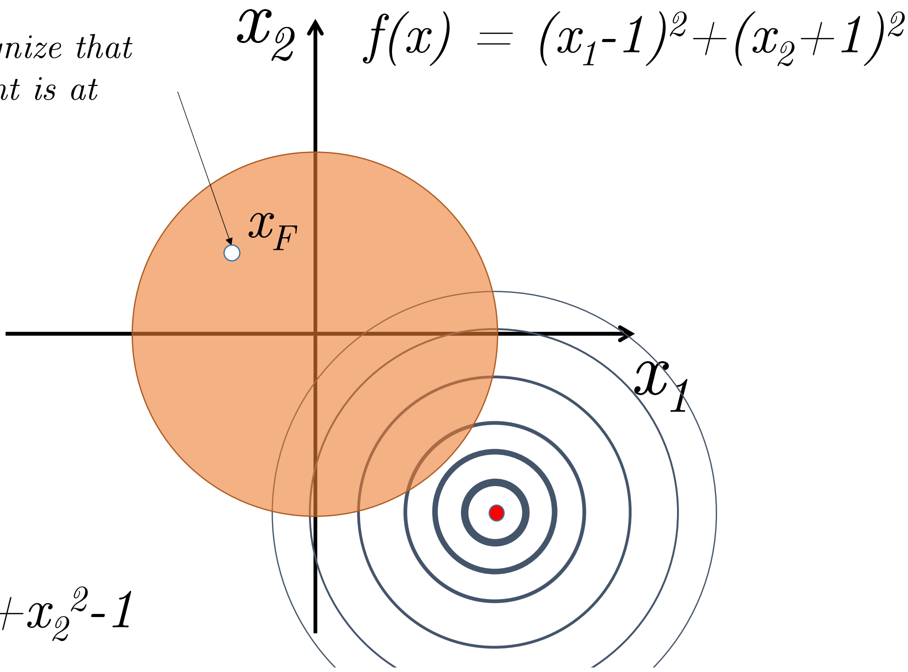
$$f(x) \to \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } g(x) \leq 0$$

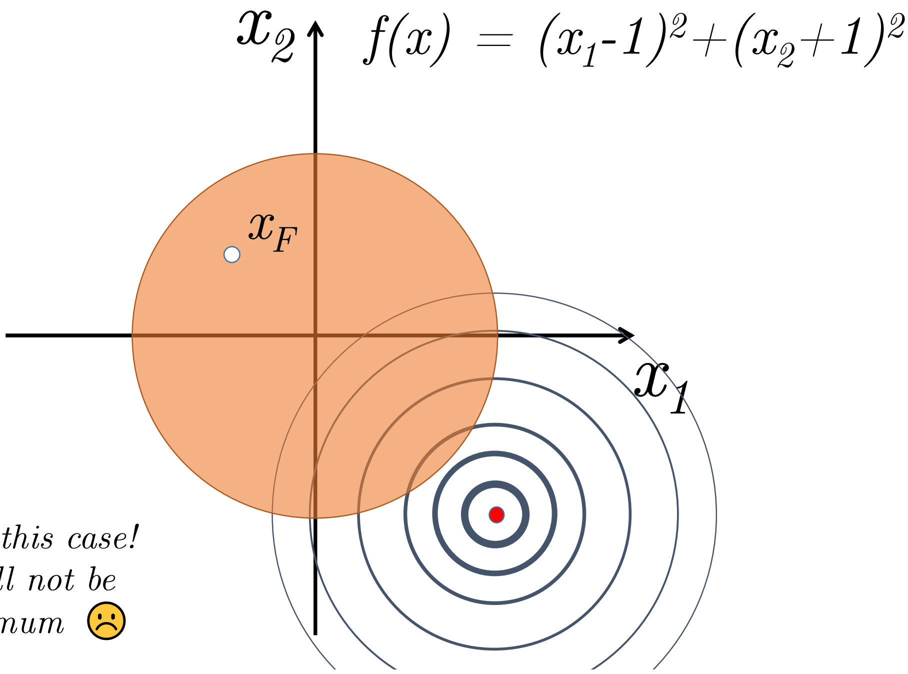




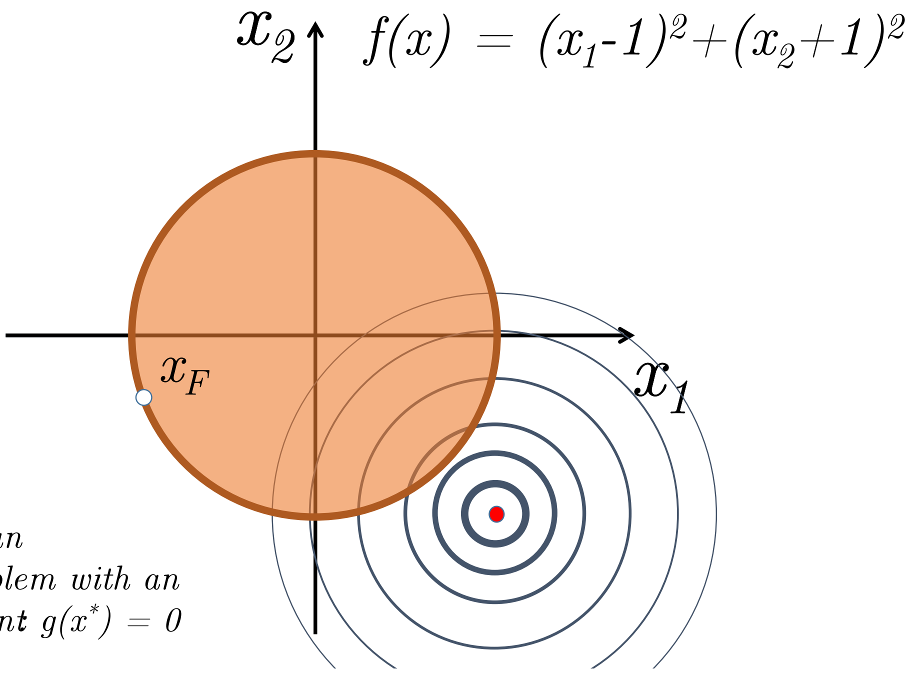
How can we recognize that some feasible point is at local minimum?



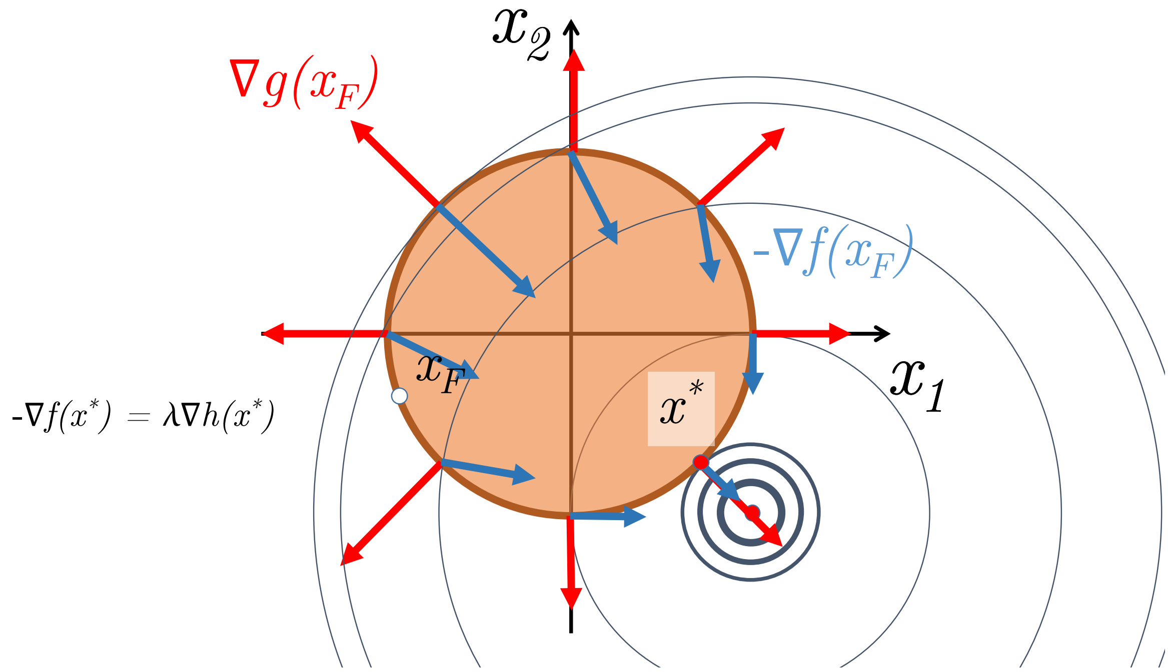
$g(x) = x_1^2+x_2^2-1$

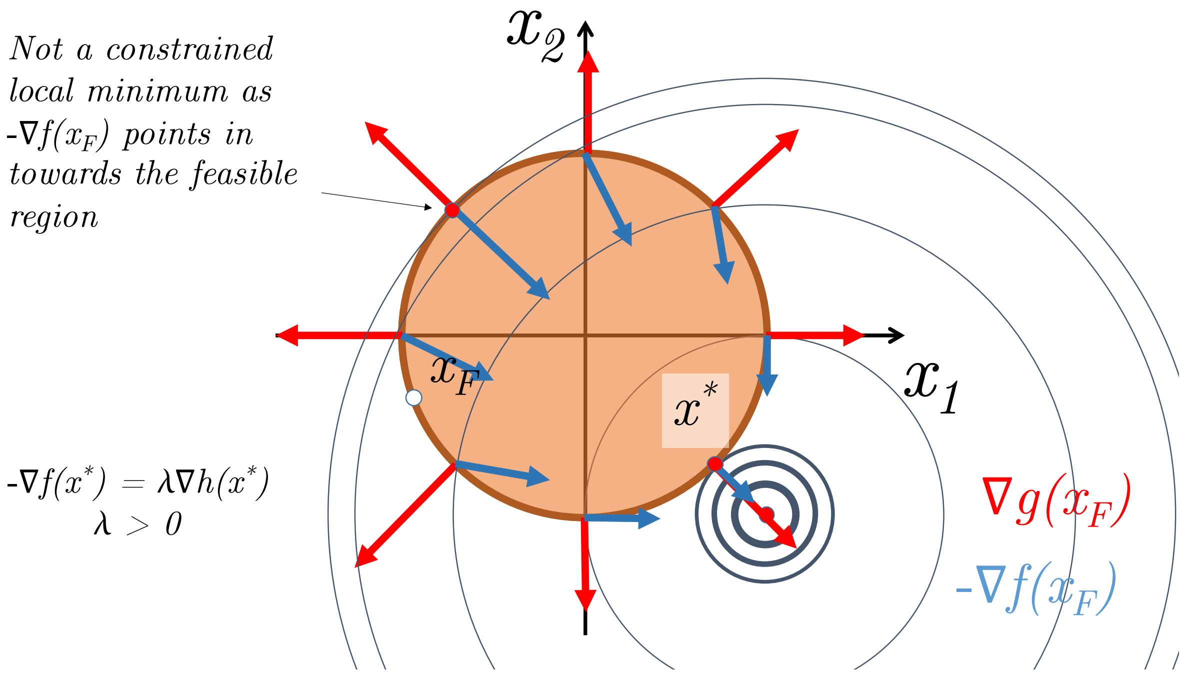


Not very easy in this case! Even gradient will not be zero at local optimum ☹



Effectively have an optimization problem with an equality constraint $g(x^*) = 0$





So, we have a problem:

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } g(x) \leq 0$$

Two possible cases: НЕ АКТИВНО АКТИВНО

$g(x) \leq 0$ **is inactive.** $g(x^*) < 0$	$g(x) \leq 0$ **is active.** $g(x^*) = 0$
$g(x^*) < 0$ $\nabla f(x^*) = 0$ $\nabla^2 f(x^*) > 0$	Necessary conditions $g(x^*) = 0$ $-\nabla f(x^*) = \lambda \nabla g(x^*), \lambda > 0$ Sufficient conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0,$ $\forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y = 0$

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{aligned}
f(x) &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } g(x) &\le 0
\end{aligned}
$$

Let's define the Lagrange function:

$$
\boxed{L(x, \lambda) = f(x) + \lambda g(x)}
$$

Then x^* point - local minimum of the problem described above, if and only if:

$-\nabla f(x^*) = \lambda^* \nabla g(x^*)$

$$
\begin{aligned}
&(1)\ \nabla_x L(x^*, \lambda^*) = 0 \\
&(2)\ \lambda^* \ge 0 \\
&(3)\ \lambda^* g(x^*) = 0 \\
&(4)\ \underline{g(x^*) \le 0} \\
&(5)\ \langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0 \\
&\forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y \le 0
\end{aligned}
$$

$g(x^*) = 0,\ \lambda^* \neq 0$ АКТИВНО

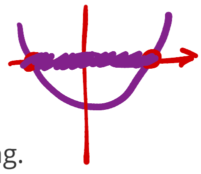
$g(x^*) < 0,\ \lambda^* = 0$ НЕ АКТИВНО

It's noticeable, that $L(x^*, \lambda^*) = f(x^*)$. Conditions $\lambda^* = 0, (1), (4)$ are the first scenario realization, and conditions $\lambda^* > 0, (1), (3)$ - the second.

General formulation

$$
\boxed{
\begin{aligned}
f_0(x) &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } f_i(x) &\le 0,\ i = 1, \ldots, m \\
h_i(x) &= 0,\ i = 1, \ldots, p
\end{aligned}
}
$$

$h_i(x) = 0$



This formulation is a general problem of mathematical programming.

The solution involves constructing a Lagrange function:

$$
\boxed{L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)}
$$

Karush-Kuhn-Tucker conditions

Necessary conditions

Let $x^*, (\lambda^*, \nu^*)$ be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*).Let also the functions f, f_i, h_i be differentiable.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda^*_i \ge 0, i = 1, \ldots, m$
- $\lambda^*_i f_i(x^*) = 0, i = 1, \ldots, m$
- $f_i(x^*) \le 0, i = 1, \ldots, m$

Some regularity conditions

These conditions are needed in order to make KKT solutions necessary conditions. Some of them even turn necessary conditions into sufficient (for example, Slater's). Moreover, if you have regularity, you can write down necessary second order conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \ge 0$ with *semi-definite* hessian of Lagrangian.

- **Slater's condition**. If for a convex problem (i.e., assuming minimization, f_0, f_i are convex and h_i are affine), there exists a point x such that $h(x) = 0$ and $f_i(x) < 0$. (Existance of strictly feasible point), than we have a zero duality gap and KKT conditions become necessary and sufficient.
- **Linearity constraint qualification** If f_i and h_i are affine functions, then no other condition is needed.
- For other examples, see wiki.

Sufficient conditions

For smooth, non-linear optimization problems, a second order sufficient condition is given as follows. The solution x^*, λ^*, ν^*, which satisfies the KKT conditions (above) is a constrained local minimum if for the Lagrangian,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

the following conditions holds:

$$\begin{aligned}
&\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle > 0 \\
&\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y \leq 0, \nabla f_j(x^*)^\top y \leq 0 \\
&i = 1, \ldots, p \quad \forall j : f_j(x^*) = 0
\end{aligned}$$

References

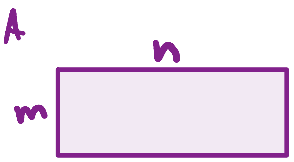
- Lecture on KKT conditions (very intuitive explanation) in course "Elements of Statistical Learning" @ KTH.
- One-line proof of KKT

Example 1

Linear Least squares Write down exact solution of the linear least squares problem:

$$\|Ax - b\|^2 \to \min_{x \in \mathbb{R}^n}, A \in \mathbb{R}^{m \times n}$$

Consider three cases:



Недоопределенная

$x + y = 5$

1. $m < n$
2. $m = n$
3. $m > n$

Решений много.
Выберем то из них, где норма минимальна

$$\|x\|^2 \to \min_{x \in \mathbb{R}^n}$$

$Ax = b \quad Ax - b = 0$

$$L(x, \lambda) = \|x\|^2 + \lambda^T (Ax - b)$$

$$\begin{aligned}
\nabla_x L &= 0 \\
\nabla_\lambda L &= 0
\end{aligned}
\quad
\begin{cases}
\underset{n\times1}{2x} + \underset{n\times m \cdot m\times1}{A^T\lambda} = 0 \\
Ax = b
\end{cases}
\quad
\begin{cases}
-\dfrac{A^T\lambda}{2} = x \\
A \cdot \left(-\dfrac{A^T\lambda}{2}\right) = b
\end{cases}$$

$$\begin{cases} x = -\dfrac{A^T\lambda}{2} \\ \underbrace{A\cdot A^T}_{\substack{m\times n \ \ n\times m \\ m\times m}}\cdot\lambda = -2b \end{cases} \Rightarrow \lambda = (AA^T)^{-1}\cdot(-2b)$$

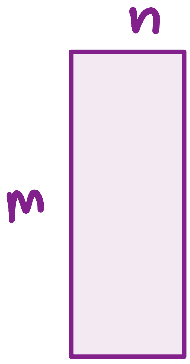
$$\boxed{x = A^T\cdot(AA^T)^{-1} b}$$

$m<n$ $\quad rg\,A = m$

$$\boxed{x^* = A^T (AA^T)^{-1} b}$$

2) $m=n$ $\qquad Ax=b \Rightarrow x = A^{-1} b$

3) $m>n$



переопределенная

$\nexists\, x^*: \ f(x^*) = 0$

$\|Ax^* - b\|^2 = 0$

$$\|Ax-b\|_2^2 \to \min_{x\in\mathbb{R}^n}$$

$$2\cdot A^T(Ax-b) = 0$$

$\underset{n\times m \ \ m\times n}{A^T A} x = A^T b$ $\qquad rg\,A = n$

$$\boxed{x^* = (A^T A)^{-1}\cdot A^T b}$$

\cross

\dagger

1) →
2) →
3) →

$$\boxed{x^* = A^\dagger b}$$

псевдообратная матрица

$$A^{\dagger} = \lim_{\alpha \to 0} \left(A^T A + \alpha \underline{I} \right)^{-1} A^T$$